

学校编码: 10384

分类号_____密级_____

学 号: 10220061150009

UDC_____

厦 门 大 学

硕 士 学 位 论 文

基于语料库的数学定义研究

A Corpus-Based Study of Mathematical Definition

卞成德

指导教师姓名: 郑泽芝 副教授

专 业 名 称: 语言学及应用语言学

论文提交日期: 2009 年 4 月

论文答辩日期: 2009 年 月

学位授予日期: 2009 年 月

答辩委员会主席: _____

评 阅 人: _____

2009 年 4 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（ ） 1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

（ ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士论文摘要库

内容摘要

定义研究不仅是术语学的一个重要方面，同时也是词典学的重要研究内容，在信息大爆炸的今天，术语定义的自动获取也日益受到了中文信息处理领域的关注。本文选取了一个特定领域的定义——数学教材中的定义作为我们的研究对象。

本文采用基于语料库的方法，以义务教育阶段的中小学数学教材中的定义为研究对象，结合定量分析和定性研究，对其中的定义从语言学角度和定义自动获取两个角度进行了探索，研究的主要内容有：

首先，本文在前人研究的基础上，对定义与概念、定义与释义的关系进行了探讨，指出了定义与释义之间的联系与不同。

其次，从认知和表述手段两个角度，对定义的方式方法进行了分类，并简要指出了各种分类方法的特点（优缺点）。

第三，在基于语料库的基础上，面向定义的自动获取，总结了数学教材中定义的模式，并将定义模式表示成规则模板，采用正则表达式匹配的方法对数学概念的定义进行了提取研究。实验证明，该提取方法效果较好，正确率和召回率分别达到了 90.09%和 99.03%。

第四，对数学概念的定义的用字情况进行了统计，并根据频次和覆盖率将数学概念的定义用字做了分级，同时对数学术语的用字位置特点进行了研究，这些研究对于术语提取以及自动分词标注都具有一定的参考价值。

关键词：定义；定义方法；定义提取；用字分析

Abstract

The research of definitions is a important aspect of terminology and also takes very important part in Lexicography. Today, along with the rapid development of information, the automated definitions extracing technique has been paid attention more and more. In this article, we chose definitions in a particular areas—mathematical definitions as our object of study.

This artical uses Corpus-based research method and chooses the mathematical definitions in math textbooks of which are used in Chinese primary and middle school as research object. In addition, we combine the quantitative analysis method with the qualitative analyses method, which guiding us to make an all-round study about the mathematical definitions. The main research contents of this article are:

First, the article summarizes the previous research about the conceptions of definitions and makes a discussion about the relationship of the definitions、paraphrases and conceptions. Thereby, we clarify the related questions especially the different between definition and paraphrase.

In the second, the article summarizes the disadvantages of previous research on methords of the definitions, and divides the definitions methords into two kinds at a different angle. One angle is thinking, the other is supplementary means. Besides, we discusses briefly about the characteristics of different classifications.

Third, the article summarizes modes of mathematical definition based on the corpus and expressed the modes as regulation templates , Then extract the mathematical definition from the corpus based on the method of regular experssion matching. The experimental results show that the proposed method is efficient ,both recall ratio and precision ratio are measure up.

At last,we make a statistics to the used in mathematical definition, according the frequency and coverage rate,we classify the characters into three levels,and make an analysis to each level. Besides, we also do a research to the characteristics of using

characters of mathematical terms, It is provided with some reference value to the term-extracting and automatic word segmentation.

Key words: definitions; method of definitions; definitions-extracting;
analysis of using characters;

厦门大学博硕士论文摘要库

目 录

第一章 绪 论.....	1
第一节 研究的起因.....	1
第二节 相关研究综述.....	2
第三节 研究对象、研究方法、研究意义.....	6
第二章 定义以及相关概念的分析.....	9
第一节 定义.....	9
第二节 定义与相关概念的关系.....	10
第三章 数学定义的定义方法.....	13
第一节 定义方法的相关研究.....	13
第二节 数学定义的定义方法.....	15
第四章 数学定义的提取研究.....	21
第一节 语料来源及定义提取的基本方法.....	21
第二节 数学定义的获取.....	24
第五章 数学定义用字分析.....	31
第一节 数学定义用字分析.....	31
第二节 数学定义用字与《现代汉语常用字表》的比较.....	36
第三节 数学定义中的术语用字分析.....	37
第六章 结 论.....	43
参考文献.....	45
后 记.....	48

Contents

Chapter 1 Introduction	1
1.1 Reason for the Reaserch	1
1.2 The General Review Related to the Studay	2
1.3.The Object 、 Significance and Method of the Study	6
Chapter 2 Analysis on the Definition and Other Concepts Related to the Definition.....	9
2.1 Definition	9
2.2 The Relationship of Definition and Other Concepts Related to the Definition.....	10
Chapter 3 General Method of Defining Mathematical Definition	13
3.1 Reaserch Related to the Method of Definition	13
3.2 Method of Defining Mathematical Definition	15
Chapter 4 Reaserch on the Extraction of Mathematical Definition.....	21
4.1 Source of the corpus and general method of definition ‘s extraction.....	21
4.2 Acquire mathematical definition	24
Chapter 5 Analysis on Mathematical Definition.....	31
5.1 Analysis on Chinese Characters Which Used in Mathematical Definition.....	31
5.2 Differences Between Mathematical Definition and 《Basic Chinese Characters of Modern Chinese》	35
5.3 Analysis on Chinese Characters of Mathematical Term	37
Chapter 6 Concluding Remarks	43
References.....	45
Acknowledgement.....	48

厦门大学博硕士论文摘要库

第一章 绪论

第一节 研究的起因

早在古希腊、罗马时代，哲学领域的学者们就开始研究各种哲学概念、自然科学概念、人文科学概念，这可以看做是术语定义研究的雏形。19 世纪末 20 世纪初，一些西方的译著开始逐渐被介绍到中国，在翻译的过程中，遇到了术语的译名问题，这时，术语的制定和规范化开始被人们所重视。随着知识的加速膨胀和科技的发展，各领域的术语越来越多的出现，人们开始逐步认识到术语研究的重要性。如何发现新的术语、如何给术语下定义、如何对现有的术语进行审定和规范，成为语言学界、术语学界、中文信息处理学界普遍关心的问题。于是，关于术语词典编纂、术语标准化、术语数据库、术语定义的研究开始展开，一门新的研究领域——术语学随之产生。

现代术语学的重要研究对象之一就是术语的定义。中国术语学的创始之作——冯志伟先生的《现代术语学引论》中，就专门列出一章介绍术语的定义、定义的分类等知识，并指出了定义研究的意义：“术语的概念是靠定义来揭示的，定义是对某个知识领域的术语进行解释和阐述。定义的科学性和逻辑性直接影响到人们对于术语的理解”^①，由此可见定义研究的重要性。近年来，关于定义的研究迅速展开，定义的本体研究、定义的系统性研究、定义的自动提取研究、定义的生成研究、法律定义研究、经济学定义研究、哲学定义研究……等等，都取得了较大的成绩。

在数学教育领域，数学定义研究历来为广大数学教育工作者所重视。因为数学知识集中体现在数学术语当中，这些术语是人类数学研究的成果，是数学知识在语言中的结晶。如何清晰的、简洁的、精确的阐释数学术语所包含的概念，使学生更容易的掌握数学知识，就涉及到数学定义各方面的研究。因此，数学定义的定义方法研究、定义特点研究、定义的认知心理学研究是数学定义研究的主要内容。

^① 冯志伟. 现代术语学引论[M]. 北京: 语文出版社, 1997. 8: 31.

本文就是在此背景下,综合吸收利用当前术语定义研究的相关成果,采用基于语料库的方法,结合定量和定性分析,面向实际应用,分别从数学定义的定义方法、定义提取模式、用字情况等各个角度,对现行义务教育阶段数学教材中出现的数学定义进行了较为详细的研究,期望研究成果能够在研究定义的提取、了解数学定义的真实存在状况以及数学定义教学等方面提供一定的参考作用。

第二节 相关研究综述

1、术语学领域的定义研究:

定义的研究一直是术语学的研究的重要内容。术语学是研究全民语言词汇中的专业术语规律的一门语言学科,主要研究术语的概念和概念系统、术语的定义、术语的标准化、术语数据库的建设、术语的语法和语义结构等。

在现代术语学产生之初,现代术语学四个学派之一的德国—奥地利学派就主张,研究术语首先要划分概念,然后才能划分概念的名称,概念系统是术语的基础。因此,术语学的研究应该先从概念出发,即先从概念的定义开始,而不是先从词语开始。“术语学应该研究概念的本质、概念的产生、概念的特性、概念之间的相互关系、概念系统的结构、概念的描述和定义等等”,并认为语言学只对语言进行描写,而术语学不仅要对语言进行描写,而且还要对语言进行规定。该学派认为定义在术语学研究中占有特殊重要的地位,为了保证术语定义的一致性,他们建立了一套严格的定义方法^①。德国—奥地利学派的这些观点是有关定义研究的最早论述,他们肯定了定义研究的重要性,强调对定义方法进行深入研究。

冯志伟在《现代术语学引论》一书中,介绍了术语下定义的一般原则和方法,阐述了内涵定义、外延定义、上下文定义、定义的规则、定义的系统性、定义的辅助手段等问题,是国内最早的关于定义的研究。另外,他也对术语形成的经济律做了深入的研究,指出了术语系统的经济指数,单词的术语构成,并把术语定

^① 冯志伟. 现代术语学引论[M]. 北京:语文出版社, 1997. 8:6.

义分为了六类，分别为：定义中的种差部分揭示了被定义事物本质特点的“实质性定义”；定义中的种差部分表示被定义事物发生、来源和形成情况的“发生性定义”；定义中的种差部分表示事物所起的作用的“功能性定义”；定义中的种差部分表示事物发生的原因的“因果性定义”；定义中的种差部分表示事物所在的位置的“空间性定义”；定义中的种差部分表示事物的成分、结构等的“结构性定义”。冯志伟先生关于定义的研究具有开创性的指导意义，但作为国内定义研究的开创者，他对定义的研究是从宏观角度进行论述的，没有更好对定义进行详细深入的分析。

刘青对定义方法以及下定义应该遵循的基本原则作了研究，认为“把术语学作为一个体系来研究，术语仅有定名仍不完备，还必须有准确的定义加以阐释”^①。他归纳了科技术语定义的几种形式，包括属加种差的方法、特殊方式揭示术语内涵的方法、数字式定义方法，并指出了定义的基本要求，包括概念准确、紧跟时代、反映本质、资料可靠；他还提出了拟定定义过程中应该注意的问题，包括定义应当确定所界定概念中最基本的特性、体现出概念在概念体系中的位置、更清楚的表明此术语的实际意义以及定义的理想结构由单句或短语构成、定义应当采用属+种差形式等 11 点原则。他的研究对于定义的拟定、术语定义的标准化起到了很大的指导作用。同冯志伟先生一样，他的研究主要面向术语标准化的宏观角度，并没有对定义进行微观细致的研究。

2、中文信息处理领域的研究

信息时代新的概念、术语层出不穷，因此，发现和提取术语以及术语的定义显得非常重要，国内的学者对定义的计算机自动提取进行了研究。

中科院自动化所张艳等人于 2003 年做了汉语术语定义的结构分析和提取方面的工作^②。他们从科技期刊的语料中选出大量定义句，面向自动提取，对这些定义的定义模式进行人工总结，并且对语料进行了分词和词性标注处理，然后应用句法分析工具分析出句子中的短语成分，并根据汉语句子的句型结构，总结出

^① 刘青.关于科技术语定义的基本问题[J].术语学研究.2004.(3):13-14

^② 张艳 宗成庆 徐波.汉语术语定义的结构分析和提取[J].中文信息学报.2003,17(6):9-16.

术语定义的结构特点,自动对其他语料库中的定义进行了提取,并且借助语义、句法等规则进行了定义的生成研究。

许勇、荀恩东等人,在贾爱平研究的基础上,利用计算机技术,自动从网上直接获取原始文本,利用定义模式(语言学规则)对定义进行了提取^①。张榕在前人研究的基础上,采用规则和统计相结合的提取方法,从真实文本中提取定义,并且她提出了“词语的定义隶属度”和“句子的定义隶属度”两个概念,由术语定义的匹配规则和排除规则筛选出候选定义,通过计算句子的定义隶属度以及向量空间模型的相似度,给出了每个定义的权重(准确度),根据权重从候选定义中选取最为准确的定义。除此之外,她还利用《知网》进行义原、义项、词语以及句子的相似度计算,实现了术语定义的按领域聚类^②。张榕的研究突破了单纯基于规则的定义提取技术,结合了统计方法,大大提高了抽取准确率,把定义自动提取技术提高到了一个新层次。但是,我们也可以看出,她的统计方法是基于规则将定义提取出来之后才使用的,目的是对提取结果应用统计方法来提高准确率,而不是在提取过程中将规则和统计的方法相结合然后输出提取结果,因此,从这一点上来说,她的定义提取技术仍然是基于规则的。

陆勇在面向信息检索的汉语同义词自动识别研究中涉及到了释义信息的定义模式^③。在他的研究中,把词典释义作为知识获取的语料,根据词典中的释义,自动进行同义词的提取和识别。他的研究采用的语料库是包含有 3800 多个经济词汇的注释语料,该语料库中的主要词汇注释来自于《中国大百科全书·经济卷》和《当代金融词典》。他提出了四种定义模式以及这四种模式的主要特征词汇:一是同义词式定义,用一个与被定义项具有相同意义的另外一个词来进行词汇的定义;二是反义词式定义,就是指词汇的定义项是由与被定义项具有相反意义的内涵来说明词汇的含义;第三种是列举式定义,即通过列举被定义项所表达的概念,从而显示被定义词语的意义的方法,根据其列举的完全与否,又可以分为:枚举式和穷举式;第四种是混合式定义模式,这类模式主要是上述模式的综合。

^① 许勇、荀恩东. 基于互联网的术语定义获取系统[J]. 中文信息学报. 2004(4):25

^② 张榕. 术语定义抽取_聚类与术语识别研究[D]. 北京语言大学博士论文, 2006.6

^③ 陆勇,侯汉清. 用于信息检索的同义词自动识别及其进展[J]. 南京农业大学学报(社会科学版), 2004,(03).

3、数学教学界关于数学定义的研究

李树臣（2000）对初中数学教材中概念的定义方式进行了研究，提出了属+种差、发生性定义和派生定义、关系定义、外延定义、否定式定义、描述性定义和公理定义、形式定义等 7 种定义方式^①。但是他的分类没有依据统一的标准，同一个定义可以归到其中的两类甚至三类当中。

辛颖（2003）认为，数学概念构成了数学知识的基础。要搞好概念的教与学，必须了解概念的定义方式及其定义规则。并把数学概念的定义方式归纳为五种，分别是：提示概念内涵、揭示概念外延、相关概念、否形式定义、序列性定义，除此之外还有公理定义方式和不定义的概念定义方式^②。

以上二人对定义方法的研究大多属于经验性的总结，缺乏理论指导，因此他们的分类没有形成系统。还有很多研究者探讨了数学定义在教学中的重要性，例如：刘勋（2007）《不可忽视的数学概念学习》、陆海泉（2008）《把握概念是学好数学的关键》等等，这些研究多是根据自己的教学经验进行散论性质的探讨，普遍缺乏理论支持。

4、其他领域的定义的研究

在各个学科领域，定义都是研究的重要内容，所以很多学者都对自己领域的定义进行了研究。例如，李鸥对经济概念及其定义的方法进行了研究，缪四平分析了法律定义的特点与作用、法律定义的基本类型，并重点考察了法律定义不同于普通定义的特殊定义形式^③。孔易人的《评哲学教材中的三种定义模式》对哲学教材中的几种错误定义进行了分析。

综观以上各个研究领域的定义研究，我们可以看出，术语学领域的定义研究注重理论，多停留在宏观层面，探讨定义的一般问题，研究没有深入。中文信息处理领域的研究基本上是面向工程，不重视定义的本体研究，定义模式和方法的研究都是服务于他们的软件系统设计。其他领域的研究大都缺乏系统性，一般根

^① 李树臣.论数学概念的分类形式[J].山东教育.2007(17):26-29

^② 辛颖.数学概念的定义方式[J].安徽电子信息职业技术学院学报.2003(1):12-13

^③ 缪四平.法律定义研究.[J].华东政法学院学报, 2003, (3):48-51

据个人经验进行总结。总体来看，对定义的系统的、量化的分析研究尚未开展。

第三节 研究对象、研究方法、研究意义

1、本文的研究对象

2001 年《全日制义务教育数学课程标准（实验）》、《普通高中数学课程标准（实验）》施行以来，各种版本的数学教材层出不穷，“迄今已有小学 6 套，中学 9 套，高中 6 套共 21 套根据《标准》要求编写的新数学教材”^①，分属 13 家出版社出版。尽管数量众多，但是从小学到高中全系列出版的并不多。人民教育出版社和北京师范大学出版社出版的中小学数学教材种类齐全，使用范围较广，可以代表了国内数学教材的一般情况。本文的研究对象是这两套教材中出现的数学术语的定义，下文我们简称它们为数学定义。这些数学定义我们利用自动提取技术提取出来，并对提取结果进行人工校对后，将结果导入数据库，然后对这些原始语料进行了多个方面的统计分析，对数学定义的各种属性进行了标注，从而形成了一个数学定义语料库。

对数学教材语言的研究，可以促进数学教育和教学，对数学定义的研究，对于教材的规范和内容的科学性都有意义。目前，学科语言的研究正在受到各方的关注。如李宇明先生在 1986 年就对数学教材语言中的数量词语和代词的语法特点、比较句的类型、祈使句和疑问句等做了深入的分析，总结出了数学语言不同于自然语言的一些特点，提出了很多新颖独到的见解^②。本文是从语言学和定义自动获取的角度对数学定义进行的研究，不涉及数学定义本身的知识内涵。

2、本文的研究方法

2.1 基于语料库的分析方法

基于语料库的研究方法是理性主义与经验主义的结合，语料库与计算机技术

^① 孙晓天. 近年来我国中小学数学教材建设述要. [J]. 数学教育学报, 2008, (4).

^② 李宇明. 数学语言初见[J]. 语文教学与研究, 1986. (3-6).

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库